Dynamics of Learning in Feature Discovery Networks

Todd K. Leen

Oregon Graduate Institute
Department of Computer Science
and Engineering
19600 N.W. von Neumann Drive
Beaverton, OR 97006-1999 USA

Technical Report No. CS/E 90-013

August, 1990

Dynamics of Learning in Feature Discovery Networks

Todd K. Leen*

Dept. of Computer Science and Engineering
Oregon Graduate Institute of Science and Technology
19600 N.W. von Neumann Drive
Beaverton OR 97006-1999

tleen@cse.ogi.edu

Aug. 1990; revised Mar. 1991

Abstract

We address the dynamics of learning in neural feature-discovery networks. The models introduced incorporate feed-forward connections modified by a Hebb law, and recurrent lateral connections modified by an anti-Hebb law.

The stability of equilibria depends on both the learning rates in the system, and the second order statistics of the ensemble of inputs. We derive conditions for stability of equilibria, and use bifurcation theory to explore the behavior near loss of stability. The bifurcation analysis uncovers previously overlooked behaviors, including equilibria that consist of mixtures of the principal eigenvectors of the input auto-correlation, as well as limit cycles. The results provide a more complete picture of adaptation in Hebbian feature-discovery networks.

*This work was supported by the Office of Naval Research under contracts N00014-88-K-0329 and N00014-90-1349 and by DARPA grant MDA 972-88-J-1004.

1 Introduction

One of the problems faced by both natural and artificial adaptive systems is the construction of efficient representations of the environment. In recent years there has been considerable interest in the notion that local Hebbian adaptation can provide a mechanism for building such representations. Numerous theoretical studies suggest, for example, that receptive field organization similar to that found in striate visual cortex can be formed adaptively, in direct response to environmental stimuli [1, 2, 3, 4].

Oja [5] made a remarkable observation that provides a link between physiologically motivated learning rules and ideas from signal processing. Oja showed that a simple model neuron, with a particular Hebbian adaptation rule, develops into a filter for the first principal component of input distribution. Linsker [6] extends this, suggesting that perceptual systems organize themselves to maximize information transfer.

Several researchers extend Oja's work, suggesting Hebbian networks that perform a complete principal component analysis (PCA). Oja and Karhunen [7] discuss an algorithm that maps to a three-layer, feed-forward network of linear neurons [8]. Sanger [9] proposes an algorithm that uses a set of cascaded feedback projections to the input space, and gives a convergence proof based on Oja's original work. This architecture singles out a particular cell for each principal component. Both of these models are robust, and useful for signal encoding applications [10].

More recently, models that use lateral signal flow to force different cells to tune to different statistical features have appeared. Foldiak [11] employs lateral connections that develop according to an anti-Hebbian rule. The cells in this algorithm do not filter for the principal components, but rather for mixtures of the principal components. We will show that this arises from a bifurcation that results from the form of the anti-Hebb rule. Rubner et al [12] propose a similar model, but with cascaded lateral connections. Like Sanger's scheme, this architecture singles out a particular cell for each principal component.

From a dynamical viewpoint, these algorithms are poorly understood. The goal of this paper is to help form a more complete picture of feature-discovery models that use lateral signal flow. We introduce two new models, with particular emphasis on their learning dynamics. Our models incorporate Hebbian and anti-Hebbian learning, and recurrent lateral connections. There

are no architecturally distinguished cells in these models. This enhances their biological plausibility. The model development results in anti-Hebbian rules that depart from the forms previously given in the literature. We give stability analyses and derive bifurcation diagrams for the models.

Stability analysis shows that the adaptation of the lateral connections has to be fast in order for the network to perform PCA. The bifurcation analyses reveal behaviors that previous researchers have overlooked. These include equilibria in which the weight vectors are combinations of the eigenvectors of the input's auto-correlation, as well as limit cycles. Since networks have high-dimensional equations of motion, we have employed a computer algebra system[1] for the bifurcation calculations.

In the next section, we make some general observations on stability that hold for a broad class of models. In section 3 we develop a model that treats the lateral connections to the lowest non-trivial order. In section 4 we treat the lateral signal flow to all orders.

2 Extending the Single-Neuron Model

Oja [5] showed that a single model neuron, with linear post-synaptic response and synaptic weights which develop under a Hebbian learning rule, develops to act as a filter for the first principal component of the input distribution. In this model the input vector, $x \in R^N$, is a random variable drawn from a stationary probability distribution. The vector of synaptic weights is denoted ω. The post-synaptic response is given by

$$y = \omega \cdot x = x^T \omega. \tag{1}$$

where the superscript 'T' denotes transpose. The Hebb rule for the adaptation of the synaptic strengths is

$$\delta\omega \;=\; \gamma[\, xy - y^2\, \omega\,] \tag{2}$$

where $\delta\omega$ is the change in the weight vector in reponse to the pattern x, and γ is the learning rate. The first term in (2) is the usual Hebbian term by which the synaptic weight changes according to the correlation between the input signal and the post-synaptic response. The second term in (2)

[1]Mathematica version 1.2, Copyright 1988 & 1989,Wolfram Research Inc.

is an active decay which prevents the magnitude of the weight vector from diverging [13, 5].

Under specific conditions on γ [5, 14, 7], the discrete update in (2) is equivalent to the ensemble-averaged differential equation,

$$\dot{\omega} = < xy > - < y^2 > \omega, \tag{3}$$

where $< \ldots >$ denotes the average over the ensemble of input patterns. Substituting (1) into (3) leaves

$$\dot{\omega} = Q\omega - (\omega \cdot Q\omega)\omega \equiv f(\omega), \tag{4}$$

where Q is the auto-correlation of the input patterns, with matrix elements $Q^{ij} = < x^i x^j >$. We denote the (unit-magnitude) eigenvectors of Q by e_i, $i = 1 \ldots N$. The corresponding eigenvalues are assumed to be ordered as $\lambda_1 > \lambda_2 > \ldots > \lambda_N$.

It is straightforward to verify that (4) has equilibria when the weight vector is along, or opposite, any of the eigenvectors of the autocorrelation. However, linear stability analysis shows that only the equilibrium at $\omega = \pm e_1$ is stable, the other equilibria are saddle points. In the basis of eigenvectors of Q, the linearization of the vector field of (4) at the equilibrium $\omega = \pm e_i$ is

$$\mathrm{Df}(\pm \mathrm{e_i}) = \begin{Bmatrix} \lambda_1 - \lambda_i & & & & & & \\ & \lambda_2 - \lambda_i & & & & \mathrm{O} & \\ & & \ddots & & & & \\ & & & -2\lambda_i & & & \\ & & & & \lambda_{i+1} - \lambda_i & & \\ & \mathrm{O} & & & & \ddots & \\ & & & & & & \lambda_N - \lambda_i \end{Bmatrix}. \tag{5}$$

The equilibria are asymptotically stable provided all of the eigenvalues of Df have negative real part. These eigenvalues can be read directly from (5). For arbitrary i, Df has $i - 1$ positive eigenvalues, corresponding to the unstable eigenspace, and $N - i + 1$ negative eigenvalues, corresponding to the stable eigenspace. Small perturbations along the directions in the stable eigenspace are exponentially damped, while those in the unstable eigenspace are amplified. Clearly, the only asymptotically stable equilibrium is at $i = 1$.

In fact it is fairly straightforward [5, 13] to show that the weight vector asymptotically approaches $\pm e_1$, the eigenvector of the auto-correlation corresponding to the largest eigenvalue. The variance of the cell's response is thus maximized, and the cell acts as a filter for the first principal component of the input distribution [5, 8].

2.1 Stability Requirements

Our goal is to extend this scheme to a system of $M < N$ cells whose weight vectors converge to the leading M eigenvectors of the auto-correlation. Alternatively, one may relax this goal, requiring that the weight vectors converge to a spanning basis for the $M-$ dimensional principal component subspace. However, the solution of eigenvectors of the auto-correlation is convenient for analytic purposes and we will begin our discussion with this case.

Before discussing particular models, we point out a general feature that will arise in a broad class of models designed to carry out our purpose. We show that the extended system will be parameterized by a coupling constant (or learning rate) that has a critical value for stability of the desired equilibrium.

Consider a set of M linear neurons with weight vectors $\omega_1, \ldots, \omega_M$ connecting each to the N-dimensional input space. We replicate (4) for each cell, adding suitable interaction terms designed to force the weight vectors to converge to different eigenvectors of the input's auto-correlation. For the purpose of this section, it is sufficient to assume that the interactions between the cells are sufficiently weak that we may treat them as a perturbation of the dynamics of (4). We assume the form,

$$\dot{\omega}_i = f_i(\omega_i) + C\, G_i(\omega_1, \ldots, \omega_M, \eta), \quad i = 1, \ldots, M \tag{6}$$

where f_i is a vector field of the same form as in (4), C is a coupling constant which specifies the strength of the interactions, G_i carries the interactions between the cells, and the η are auxiliary variables appearing in the interactions. In general, there will be additional differential equations describing the evolution of the auxiliary variables. We write these as

$$\dot{\eta}_{ij} = f_{\eta_{ij}}(\omega, \eta, C), \tag{7}$$

where $f_{\eta_{ij}}$ contains terms which describe the interactions between the weight vectors and the auxiliary variables. In the following sections, the auxiliary

variables will take the form of a set of lateral couplings between the cells, and the constant C will appear as the learning rate associated with these lateral couplings.

Consider the behavior at $C = 0$. We assume that the system (6, 7) has an equilibrium at $\omega_i = \omega_i^0 = e_i, \quad \eta_{ij} = \eta^0{}_{ij}$. This equilibrium cannot be stable. To see this, we observe that at $C = 0$, (6) is decoupled from (7). The equilibrium $\omega_i = e_i$ has unstable eigenspaces for each $i > 1$, so the system (6) and (7) will have unstable eigenspaces at the equilibrium in question.

Assuming that the linearization of (6, 7) has no eigenvalues with zero real part at $C = 0$, the equilibrium will persist for small values of C and the unstable directions will remain unstable [15]. Thus there is a minimal value for $|C|$, below which the equilibrium is unstable.

3 Minimal Coupling

In this section we introduce interactions between the cells in a minimal fashion. This model is obtained by writing a potential for the single-neuron model of §2, and augmenting this single-neuron potential with interaction terms designed to orthogonalize the weight vectors. These interactions are made local by introducing lateral couplings between the cells. The derivation leads naturally to an anti-Hebbian adaptation rule for the lateral connections.

3.1 Potential Formulation

It is widely recognized that the model in (4) performs hill-climbing on the variance of the cell response. The Hebb term in (4) is recovered by taking the gradient of

$$U(\omega) = -\frac{1}{2} < y^2 >= -\frac{1}{2}\, \omega \cdot Q\, \omega$$

with respect to ω. We extend the system to an array of M cells, with weight vectors ω_i, $i = 1, \ldots, M$, and linear responses $y_i = \omega_i \cdot x$. The potential for the array is

$$U(\omega) = -\frac{1}{2} \sum_{i=1}^{M} \omega_i \cdot Q\omega_i. \tag{8}$$

Following Yuille et al. [16] (see also [17]) we introduce an interaction potential that penalizes correlations between the cell responses

$$I = \sum_{i \neq j} \frac{1}{2} (< y_i \, y_j >)^2 = \sum_{i \neq j} \frac{1}{2} (\omega_i \cdot Q \, \omega_j)^2. \tag{9}$$

This form elevates the potential in regions of the weight space where the cell responses are correlated or anti-correlated. We obtain the total potential by combining (8) and (9),

$$\begin{aligned} U_{Tot} &= U(\omega) + C \, I \\ &= -\frac{1}{2} \sum_{i=1}^{M} (\omega_i \cdot Q \, \omega_i) + \frac{C}{2} \sum_{i \neq j} (\omega_i \cdot Q \, \omega_j)^2, \end{aligned} \tag{10}$$

where C is a coupling constant (c.f. §2.1).

The equation of motion for the i^{th} weight vector is obtained by combining the gradient of (10) with a decay term of the form given in (4),

$$\begin{aligned} \dot{\omega}_i &= -\nabla_{\omega_i} U - (\omega_i \cdot Q \, \omega_i) \, \omega_i \\ &= Q \, \omega_i - C \sum_{j \neq i} (\omega_i \cdot Q \, \omega_j) \, Q \, \omega_j - (\omega_i \cdot Q \, \omega_i) \, \omega_i. \end{aligned} \tag{11}$$

It is helpful to rewrite (11) in terms of ensemble averages over the input patterns, and the response of the i^{th} cell, $y_i = \omega_i \cdot x$. Thus

$$\begin{aligned} \dot{\omega}_i &= < x \, y_i > - C \sum_{j \neq i} < y_i \, y_j > < x \, y_j > - < y_i^2 > \omega_i \\ &= < x \, [y_i - C \sum_{j \neq i} < y_i \, y_j > y_j] > - < y_i^2 > \omega_i. \end{aligned} \tag{12}$$

The first term on the right-hand side of (12) drives changes in the weight vector according to the correlation between the the input signal x and the modified response,

$$\tilde{y}_i \equiv y_i - C \sum_{j \neq i} < y_i \, y_j > y_j. \tag{13}$$

The j^{th} term of the sum in (13) is the response y_j gated by a factor proportional to the correlation between cells j and i. This correlation is *not* computed locally within the network. Furthermore the response of the j^{th} cell is not locally available to cell i.

3.2 Local Realization

The most natural way to localize these computations is to provide a set of lateral connections between the cells. In this section we treat these lateral signals to the lowest non-trivial order. We will continue to calculate the cell activities from the signals carried on the weights ω as in (1). In this approximation the lateral connections mediate synaptic plasticity without directly influencing the cell response.

We introduce a symmetric matrix of $M(M-1)/2$ distinct lateral connection strengths,

$$\eta_{ij}, \;\; i,j \;=\; 1,\ldots,M; \;\; i \neq j,$$

and require that these equilibriate to $-C < y_i\, y_j >$. The simplest dynamics to carry out this relaxation is

$$\dot{\eta}_{ij} \;=\; -d\,(\,\eta_{ij} \,+ C\,\omega_i \cdot Q\,\omega_j) \tag{14}$$

where d is a rate constant. This form captures the notion that the lateral connections develop to oppose correlations between the cell responses.

We make the substitution $C\,\omega_i \cdot Q\,\omega_j \rightarrow -\eta_{ij}$ in (11) to obtain the final equations of motion for the forward weights,

$$\dot{\omega}_i \;=\; Q\,\omega_i + \sum_{j \neq i} \eta_{ij}\, Q\,\omega_j - (\omega_i \cdot Q\,\omega_i)\,\omega_i. \tag{15}$$

Equations (14) and (15) are the adaptation dynamics for the system of M cells with weights from the input space ω and lateral connections η.

3.3 Stability and Bifurcation Behavior

We show that the stability of the desired equilibrium is dependent on the free parameters, in accord with the discussion in §2.1. The primary result is that the adaptation of the lateral connections needs to be *fast* relative to the adaptation of the forward weights in order for the system to perform a principal components analysis. Beyond the stability analysis, we derive bifurcation diagrams for the system.

3.3.1 Stability

By inspection (14) and (15) have an equilibrium at

$$X_0 \equiv \{\omega_i = e_i,\ i = 1, \ldots, M;\ \ \eta_{ij} = 0\} \quad \forall\, C. \tag{16}$$

To treat the stability of this equilibrium it is convenient to expand the ω_i in the basis of eigenvectors of Q, writing the components as

$$\omega_{ij} \ \equiv \ \omega_i \cdot e_j.$$

Next we collect all variables into a single coordinate vector and arrange the components as,

$$\begin{aligned} X \ = \ & [(\omega_{12}, \omega_{21}, \eta_{12}), (\omega_{13}, \omega_{31}, \eta_{13}), \ldots (\omega_{M-1,M}, \omega_{M,M-1}, \eta_{M-1,M}) \\ & \{\omega_{11}, \omega_{22}, \ldots, \omega_{MM}\}, (\omega_{1,M+1}, \ldots, \omega_{1,N}, \ldots, \omega_{MN})]. \end{aligned} \tag{17}$$

This vector contains $M(M-1)/2$ triplets of the form $(\omega_{ij}, \omega_{ji}, \eta_{ij})$, M components ω_{ii}, and $M(N-M)$ components in the last block. We write the equations of motion (14, 15) in short-hand as

$$\dot{X} = F(X) \tag{18}$$

with $F(X)$ defined by the components of the right-hand sides of (14) and (15) arranged in the same order as the coordinate vector (17). In this notation, the equilibrium (16) is at

$$X_0 = [(0,0,0), \ldots, (0,0,0), \{1,1,\ldots\}, (0,0,\ldots)] \tag{19}$$

At X_0 the linear part of the vector field takes the block-diagonal form (see the appendix)

$$\mathrm{DF_0} \equiv \left(\frac{\partial \mathrm{F}}{\partial \mathrm{X}}\right)\Bigg|_{\mathrm{X_0}} = \left\{ \begin{array}{ccccc} [\mathcal{M}_{12}] & & & & \\ & [\mathcal{M}_{13}] & & & \\ & & \ddots & & \\ & & & \{A\} & \\ & & & & (B) \end{array} \right\}, \tag{20}$$

where the 3×3 sub-blocks, $\mathcal{M}_{ij}$, $i < j$, are of the form

$$\mathcal{M}_{ij} \ = \ \left[\begin{array}{ccc} \lambda_j - \lambda_i & 0 & \lambda_j \\ 0 & \lambda_i - \lambda_j & \lambda_i \\ -C\,d\,\lambda_j & -C\,d\,\lambda_i & -d \end{array} \right], \tag{21}$$

and $\{A\}$ and (B) are diagonal matrices of the form

$$A = \begin{Bmatrix} -2\lambda_1 & & & \\ & -2\lambda_2 & & \\ & & \ddots & \\ & & & -2\lambda_M \end{Bmatrix} \tag{22}$$

and

$$B = \begin{pmatrix} \lambda_{M+1} - \lambda_1 & & & \\ & \lambda_{M+2} - \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_N - \lambda_M \end{pmatrix}. \tag{23}$$

Concentrate for the moment on the lower two blocks A and B. Since the eigenvalues of Q are ordered according to $\lambda_1 > \lambda_2 \ldots$, the elements in these blocks are all negative. Thus the invariant subspaces corresponding to these blocks are stable eigenspaces. The block B is of particular interest. This block corresponds to perturbations out of the principal component subspace. These perturbations are always damped, indicating that the space spanned by the principal eigenvectors is asymptotically stable.

The only possible instabilities arise in the 3×3 sub-blocks $\mathcal{M}_{ij}$. These blocks define invariant subspaces of DF_0, and each can be considered separately. Thus the stability problem for the entire recurrent network reduces to the consideration of a set of 3×3 matrices.

Applying the Routh-Hurwitz conditions to the characteristic equation for $\mathcal{M}_{ij}$, we find that the equilibrium X_0 is asymptotically stable provided

$$d > d_{ij_0} \equiv \frac{(\lambda_i - \lambda_j)^2 (\lambda_i + \lambda_j)}{\lambda_i^2 + \lambda_j^2} \tag{24}$$

$$C > C_{ij_0} \equiv \frac{1}{\lambda_i + \lambda_j}. \tag{25}$$

These conditions must be satisfied for all choices of the indices (i, j). The critical values C_0 and d_0 depend on the eigenvalue spectrum of Q, so the stability of the equilibrium is dependent on the second order statistics of the input signal. Note that (25) is apt to be violated for networks with a large number of cells since C_0 increases with decreasing (λ_i, λ_j).

3.3.2 Bifurcation Behavior

We want to locate the equilibria,

$$F(X(C), C) = 0 \tag{26}$$

near (X_0, C_0) and determine their stability. A direct solution of (26) is intractable. Instead we use the Liapunov-Schmidt reduction [18] to replace the high-dimensional system (26) with a low-dimensional system

$$g(z, C) = 0 \tag{27}$$

which is easily solved. The reduced system (27) is equivalent to (26) in the sense that the zeroes of g are in one-one correspondence with the zeroes of F, and the stability of the bifurcating equilibria can be inferred from g. In this sense the reduced function completely characterizes the bifurcation. The reduction is accomplished by means of a perturbation expansion about the bifurcation point (X_0, C_0). Details are given in the appendix.

We assume that the stability condition (25) is violated for a *single* pair of indices (i, j). At $C = C_{ij_0}$, $\mathcal{M}$ has a simple zero eigenvalue. We denote the corresponding eigenvector of DF_0 by v_r. In this case the reduced function is a real-valued function of the scalar variables z and C, where z is the displacement from X_0 along v_r. The equilibrium X_0 corresponds to $z = 0$.

The perturbation expansion shows that, to third order in z, the reduced function is equivalent to

$$g(z, C) \;=\; -d(\lambda_i + \lambda_j)(C - C_0)\, z + dz^3 \;+\; \ldots \tag{28}$$

which is the normal form for a super-critical pitchfork bifurcation. The bifurcation diagram (the solution set of $g(z, C) = 0$) is shown in the upper portion of Fig. 1. The branch corresponding to the equilibrium X_0 is stable for $C > C_0$ and unstable for $C < C_0$. Two *unstable* branches are present for $C > C_0$. At the equilibria on the unstable branches the forward weight vectors are mixtures of e_i and e_j, and the lateral connection η_{ij} is non-zero. The form of this bifurcation is independent of both the number of nodes M in the network, and the dimension N of the input space.

The position of stable equilibria away from (X_0, C_0) can be inferred from terms in the bifurcation expansion of order z^5 and higher, or alternatively as follows. For simplicity, consider the case of two cells. We examine the

degenerate solution for which both weight vectors are proportional to the principal eigenvector,

$$X_d \equiv \{\,(\omega,\eta) \mid \omega_1 = \pm\omega_2 = \frac{1}{\sqrt{1+C\lambda_1}}\, e_1,\ \ \eta_{12} \ = \ \mp\frac{C\lambda_1}{1+C\lambda_1}\}.$$

This is asymptotically stable provided

$$C < C_d \equiv \min\left\{ \begin{array}{c} (\lambda_1-\lambda_2)/(2\lambda_1\lambda_2) \\ 1/\lambda_1 \end{array} \right\}. \tag{29}$$

If the first condition in (29) is violated, then there is a supercritical pitchfork bifurcation, bottom portion of Fig. 1. (We have not determined the form of the bifurcation under violation of the second condition in (29)). The equilibria on the bifurcating, *stable* branches are mixtures of e_1 and e_2 with non-zero η_{12}. These branches presumably join the unstable supercritical branches of the bifurcation at $(X_0,\, C_0)$.

Numerical integration of (14) and (15) confirms this picture. For large C the M weight vectors converge to the leading M eigenvectors of the autocorrelation. For $C < C_0$, the weight vectors converge to mixtures of the leading eigenvectors. For $C < C_d$, the weight vectors collapse to the leading eigenvector. This scheme thus requires strong coupling (large C) between the cross-correlations $< y_i y_j >$ and the lateral connection strengths η_{ij} in order to effectively separate the forward weight vectors.

The insets in Fig. 1 show the receptive fields (ω_1 and ω_2) corresponding to the stable branches of the bifurcation diagram for a network of two cells. The plots show the magnitude of each of the components of ω. These configurations were generated by a correlation matrix corresponding to a 19-dimensional noise vector with short-range correlations between the components.

We confirmed the complete bifurcation diagram by numerical integration, sweeping the coupling strength up and back down through the bifurcation points. Figure 2 shows the cosine of the angle between the two weight vectors as a function of the coupling strength C. At the lowest values of C the weight vectors are opposite one-another, corresponding to X_d. As C is increased this configuration becomes unstable and the angle between the weight vectors begins to close. At the highest values of C, the weight vectors are orthogonal, corresponding to X_0.

For networks with more than two cells, there are presumably additional bifurcations along the branches emanating from (X_d, C_d). Simulations show various mixed states. Figure 3 shows receptive field configurations generated by a 3-cell model. The critical coupling value for these simulations is $C_0 = 0.294$ for these simulations. Figure (3a) shows the receptive fields for $C = 1.0$. These are the eigenfunctions of the input correlation. Figure (3b) shows the receptive fields generated at $C = 0.28$. One of the nodes has converged to the leading eigenfunction, while the other two nodes have converged to mixtures of the second and third correlation eigenfunctions. Figure (3c) shows the receptive fields generated at $C = 0.18$. Two of the nodes have converged to the principal correlation eigenfunction, while the third has converged to the second eigenfunction.

Finally if the condition on d in (24) is violated for a single pair of indices (i, j), while (25) satisfied, then the equilibrium X_0 loses stability through a Hopf bifurcation. A pair of complex-conjugate eigenvalues of DF cross the imaginary axis at

$$\pm i\Lambda_0 = \pm i\,(\lambda_i - \lambda_j)\sqrt{C\,(\lambda_i + \lambda_j) - 1}$$

and the forward weights and lateral connection strengths will begin to oscillate.

3.4 Activity-Dependent Adaptation

The conditions on the stability of X_0 suggest that the adaptation rule for the lateral connection strengths can be modified to provide a more robust system. Examining (24) and (25), it is clear that the critical value for $C\ d$ is bounded above by unity. Furthermore the relaxation rate required by (24) is bounded above by

$$\overline{d_0} \equiv \lambda_i + \lambda_j,$$

which is the sum of the node response variances at the equilibrium point X_0.

This suggests that the stability may be improved by weighting the relaxation rate of the lateral connections by the variance of the cell responses. We make this change, rewriting (14) as

$$\begin{aligned} \dot{\eta}_{ij} &= -< y_i^2 + y_j^2 >\, \eta_{ij} \;-\; C \;< y_i y_j > \\ &= -(\omega_i \cdot Q\,\omega_i + \omega_j \cdot Q\,\omega_j)\,\eta_{ij} - C\omega_i \cdot Q\,\omega_j. \end{aligned} \tag{30}$$

With this change, the critical 3×3 sub-blocks of DF_0 become

$$\mathcal{M}_{ij} = \begin{bmatrix} \lambda_j - \lambda_i & 0 & \lambda_j \\ 0 & \lambda_i - \lambda_j & \lambda_i \\ -C\,\lambda_j & -C\,\lambda_i & -(\lambda_i + \lambda_j) \end{bmatrix}, \tag{31}$$

and the conditions for stability of the equilibrium X_0 reduce to

$$C > 1, \tag{32}$$

which no longer depends on the spectrum of the auto-correlation.

4 Complete Coupling

The model presented in the previous section deals with the lateral connections only to the lowest non-trivial order in η. In this section we develop a model that takes full account of the lateral connections. We derive stability conditions for the desired equilibrium and treat the bifurcation under loss of stability. We also suggest an enhancement, similar to that in §3.4, that provides a more robust algorithm.

4.1 Equations of Motion

As in the previous section, the model consists of an array of M linear neurons connected to the $N-$dimensional input space. The array is self-connected with a set of symmetric lateral connections. Both sets of connection strengths develop dynamically.

The notation for this section departs slightly from that used previously. We define an $M \times N$ matrix of forward weights, ω, connecting the input space to the cell array. The i^{th} row of ω is the weight vector, ω_i, to the i^{th} cell of the array. The matrix η has the same structure as in the previous section. In addition, we define the vector of cell responses $y \in R^M$.

The lateral connections converging on a cell are assumed to carry signals which contribute to the cell's response in the usual fashion. The response of the cell is given by the sum of the forward-propagated signals and the laterally-propagated signals. Thus the cell responses are given by

$$y = \omega x + \eta y,$$

where $x \in R^N$ is the input pattern vector. This expression is solved for y to recover

$$y = u\,\omega\,x \tag{33}$$

where

$$u \equiv (1-\eta)^{-1} \tag{34}$$

and 1 denotes the identity matrix.

In a system with explicit node dynamics, the matrix inversion in (34) would be implicitly calculated through the *node* dynamics (assuming convergent activation dynamics [19][1])). For implementation in a digital system, or for simulation without explicit node dynamics, some form of direct matrix inversion would need to be calculated. Alternately a truncated series expansion of $u(\eta)$ seems to be a viable alternative.

The ensemble-averaged, continuous time form of the adaptation rule for the forward weights takes the form

$$\begin{aligned} \dot{\omega} &= < y\,x^T > - Diag < y\,y^T > \omega \\ &= u\,\omega\,Q - Diag\,(u\,\omega\,Q\,\omega^T\,u^T)\,\omega \end{aligned} \tag{35}$$

where *Diag* is an operator which takes the diagonal elements of its argument. The lateral connection strengths develop according to

$$\begin{aligned} \dot{\eta}_{ij} &= d\,\eta_{ij} - C < y\,y^T >_{ij} \\ &= d\,\eta_{ij} - C\,(u\,\omega\,Q\,\omega^T\,u^T)_{ij}, \quad 1 \le i \ne j \le M. \end{aligned} \tag{36}$$

The adaptation dynamics in (35) and (36) are close analogs to those of (15) and (14) in §3.2. The difference here is that the signals carried by the lateral connections are treated to all orders. This system differs from that given by Foldiak [11] by the linear term in (36).

The system has an equilibrium at

$$X_0 \equiv \{\omega_i = e_i,\, i = 1, \ldots, M;\;\; \eta_{ij} = 0\} \;\; \forall\; C. \tag{37}$$

To see this, note that at $\eta = 0$, the matrix u reduces to the identity and (35) reduces to M copies of (4). The last term of (36) reduces to

$$C\,(u\,\omega\,Q\,\omega^T\,u^T)_{ij}\Big|_{X_0} = C\,(e_i \cdot Q\,e_j) = 0, \;\; i \ne j.$$

[1] For example the system $dy/dt = \frac{1}{\tau}(-y + \omega x + \eta y)$ has a globally attracting fixed point at $y = u\omega x$ provided the matrix $(1-\eta)$ is positive definite.

having used the definition of the equilibrium point (37) and the orthogonality of the eigenvectors of Q.

4.2 Stability and Bifurcation

In order to address the stability of the equilibrium we follow the treatment of §3.3 and expand the rows of ω in the basis of eigenvectors of Q, writing the components as

$$\omega_{ij} \equiv \omega_i \cdot e_j.$$

We regroup the components as in (17) and write the equations of motion as in (18).

To perform the stability and bifurcation calculations, we expand u as a power series in η

$$u \simeq 1 + \eta + \eta^2 + \eta^3 + \dots . \tag{38}$$

The first order term is sufficient for the stability calculation. The terms through order η^3 are required to address the bifurcation.

As in §3.3.1, the linear part of the vector field at the equilibrium breaks into block diagonal form with any instabilities constrained to 3×3 sub-blocks. For the present model, these critical sub-blocks take the form

$$\mathcal{M}_{ij} = \begin{bmatrix} \lambda_j - \lambda_i & 0 & \lambda_j \\ 0 & \lambda_i - \lambda_j & \lambda_i \\ -C\lambda_j & -C\lambda_i & d - C(\lambda_i + \lambda_j) \end{bmatrix}. \tag{39}$$

The stability conditions read

$$d > 0 \tag{40}$$

$$C > C_0 \equiv \frac{d}{(\lambda_i + \lambda_j)} + \frac{(\lambda_i - \lambda_j)^2}{(\lambda_i^2 + \lambda_j^2)}. \tag{41}$$

We digress briefly to discuss the form of the adaptation rule for η (36) in relation to (40). Previous authors [11, 12, 20] advocate the use of the naive anti-Hebbian rule

$$\dot{\eta}_{ij} = - < y_i y_j >, \tag{42}$$

assuming that it is sufficient for feature extraction and clustering algorithms.

Our development shows that the naive form can be inadequate when viewed in the context of a complete system. The discussion in §3.2 leading

up to (14) shows that the departure from the naive form is quite natural. Furthermore, (40) shows that removing the term linear in η can lead to an instability.

Let us return to the model in (35) and (36). If the stability condition on C (41) is violated, then the network undergoes a Hopf bifurcation. To show this, calculate the characteristic polynomial of $\mathcal{M}_{ij}$,

$$\begin{aligned} P(L) = L^3 &+ (C(\lambda_i + \lambda_j) - d)\, L^2 \\ &+ [C(\lambda_i^2 + \lambda_j^2) - (\lambda_i - \lambda_j)^2]\, L + d(\lambda_i - \lambda_j)^2. \end{aligned} \tag{43}$$

At $C = C_0$, the roots of (43) are

$$L_0 = \frac{-(\lambda_i - \lambda_j)^2\,(\lambda_i + \lambda_j)}{\lambda_i^2 + \lambda_j^2} \tag{44}$$

$$L_\pm = \pm i \sqrt{\frac{d\,(\lambda_i^2 + \lambda_j^2)}{\lambda_i + \lambda_j}} \equiv \pm i\,\Lambda_0. \tag{45}$$

The first (44) is negative, corresponding to a stable perturbation direction. The pair of roots in (45) are pure imaginary provided $d > 0$, so DF_0 develops a pair of pure imaginary eigenvalues at $C = C_0$.

The conditions for a non-degenerate Hopf bifurcation are satisfied [18, 21]. If (41) is violated for a *single* pair of indices (i, j) then DF_0 has only a single pair of eigenvalues on the imaginary axis. Second, these eigenvalues cross the imaginary axis with non-zero speed as C passes through C_0. To verify the crossing condition we calculate the rate of change of the real part of the complex-conjugate eigenvalues at C_0. This is given by

$$\mathrm{Re}\left[\frac{\mathrm{d\,L_\pm(C_0)}}{\mathrm{d\,C}}\right] = \mathrm{Re}\left[-\frac{\partial \mathrm{P}/\partial \mathrm{C}}{\partial \mathrm{P}/\partial \mathrm{L}}\bigg|_{\mathrm{L_\pm, C_0}}\right] = \frac{-\mathrm{d}\Lambda_0^2(\lambda_\mathrm{i}^2 + \lambda_\mathrm{j}^2)(\Lambda_0^2 + (\lambda_\mathrm{i} - \lambda_\mathrm{j})^2)}{2\,[\mathrm{d}^2(\lambda_\mathrm{i} - \lambda_\mathrm{j})^4 + \Lambda_0^6]} < 0 \tag{46}$$

which confirms that X_0 is stable for $C > C_0$. Thus the conditions for a non-degenerate Hopf bifurcation are satisfied, and a one-parameter family of periodic solutions to (35) and (36) appear as C is varied in the vicinity of C_0.

We applied the technique given in [18] to determine whether the bifurcation is super, or sub-critical. The calculations were carried out with a computer algebra package and the results corroborated by independent numerical analysis at specific values of (λ_i, λ_j) as well as by simulation. We

carried out the calculations for the case of 2 cells in a 2-dimensional input space (five degrees of freedom).

We find that the direction of the bifurcation depends on the eigenvalues λ_i and λ_j appearing in $\mathcal{M}_{ij}$. The expression for the function that determines the direction of the bifurcation is exceedingly complex, and the results are best displayed pictorially. The results are presented graphically in Fig. 4. The shaded region in the plot corresponds to values of (λ_1, λ_2) for which the bifurcation is supercritical, with the periodic orbits at $C > C_0$ unstable. The unfilled region corresponds to sub-critical bifurcation with stable periodic orbits at $C < C_0$.

Simulations show that even for values of (λ_1, λ_2) corresponding to a super-critical bifurcation, there are stable periodic orbits. Figure 5 shows a series of simulations in the super-critical regime. The critical value of the bifurcation parameter is $C_0 = 0.312$. Fig. 5a shows stable oscillations for $C = 0.30$. Figures 5b and 5c were both generated from simulations at $C = 0.33$. These two plots show that a stable periodic orbit (5b) and the stable fixed point X_0 (5c) coexist at this value of C. Figure 5d shows the convergence to X_0 at $C = 0.5$. No periodic solutions were found at this value of the coupling. These simulations suggest that the complete bifurcation diagram in the super-critical regime is shaped like the bottom of a wine bottle, only the indentation of which is shown in Fig. 4.

4.3 Activity-Dependent Adaptation

As in §3.4, we can improve the stability of the present model by modulating the adaptation rate of the lateral connections according to the cell response variances. We introduce this change, replacing (36) with

$$\begin{aligned} \dot{\eta}_{ij} &= < y_i^2 + y_j^2 > \eta_{ij} \; - \; C \; < y\, y^T >_{ij} \\ &= [(u\,\omega\,Q\,\omega^T\; u^T)_{ii} + (u\,\omega\,Q\,\omega^T\; u^T)_{jj}]\; \eta_{ij} \\ &\quad -C\,(u\,\omega\,Q\,\omega^T\; u^T)_{ij}, \quad 1 \leq i \neq j \leq M. \end{aligned} \tag{47}$$

With this change, the critical sub-blocks of DF_0 take the form

$$\mathcal{M}_{ij} = \begin{bmatrix} \lambda_j - \lambda_i & 0 & \lambda_j \\ 0 & \lambda_i - \lambda_j & \lambda_i \\ -C\lambda_j & -C\lambda_i & (1-C)(\lambda_i + \lambda_j) \end{bmatrix} \tag{48}$$

and the conditions for stability of the equilibrium X_0 now read

$$C > 1 \tag{49}$$

$$C > 1 + \frac{(\lambda_i - \lambda_j)^2}{\lambda_i^2 + \lambda_j^2} \tag{50}$$

the last of which is bounded above by 2.

Lastly it is desirable to avoid calculating the matrix inversion $(1 - \eta)^{-1}$ that appears in the equation for the output activation (33). Simulations approximating this inverse by the first two terms of its series expansion

$$u \approx 1 + \eta$$

provides reasonably good convergence.

5 Discussion

We have introduced two neural models for linear feature discovery that are based on a combination of Hebbian learning and recurrent lateral connections that develop according to an anti-Hebbian learning rule. Both models employ anti-Hebbian learning rules that include a term linear in the lateral connections, thus departing from forms previously given in the literature.

The minimal model in §3 treats the lateral connections in the lowest possible order. The signals carried on the lateral connections affect plasticity but *not* the target cell activation. This is advantageous for implementation. In a model that uses lateral signal flow to affect cell activation, the values of the cell activities would require several cycles to equilibriate. Here we calculate the activities from the forward signals alone, and this can be done in a single machine cycle.

Both models have equilibria at which the network performs PCA. There is a critical coupling strength below which this equilibrium is unstable. The bifurcation analyses and simulations show that both models have ranges of the coupling, C, that support several solutions. The minimal model has stable secondary equilibria in which the forward weight vectors are combinations of the eigenvectors of the input auto-correlation. The complete model has solutions in which the forward weight vectors are oscillating combinations of the eigenvectors of the input auto-correlation. For the complete model to be

useful in this regime, learning would have to be turned off after an initial period of adaptation.

Both models expand on earlier work on Hebbian feature discovery and principal component analysis. In the limit of fast relaxation of the lateral connections, the minimal model reverts to the model given by Yuille et al. [16] for the formation of cortical simple cells. Rubner et al. [12] discuss a model with cascaded lateral signal paths; the i^{th} cell receives lateral signals from all cells j with $j < i$. The models discussed here have *full* lateral connectivity, which is more consistent with neuroscience.

Foldiak [11] discusses an algorithm with full lateral connectivity. The adaptation rule that he uses for the lateral connections has no linear term, in contrast with our anti-Hebb rule. As discused in §4, removing the linear term can result in an instability, so Foldiak's model operates at a bifurcation point of our scheme. For a system of two cells, the reduced function at this bifurcation is of the form $g(z, C) = a\ C\ z\ +\ b\ z^n,\ \ n > 7$. Small perturbations in the equations of motion, arising for example from imperfections in the physical realization, can introduce terms of lower order in z. These lower-order terms will dominate the form of the bifurcation and could radically change the location of equilibria. Thus, in the absence of the linear term, the postion of equilibria could be dominated by imperfections in the physical realization. In this sense, the model without the linear term is likely to be an incomplete specification of any physical (e.g. biological or analog VLSI) implementation.

This study suggests several areas for further inquiry. The bifurcation from the degenerate solution in §3.3.2 was explored analytically only for the case of two cells. For networks with more cells, there are presumably additional bifurcations along the branches emanating from the degenerate solution $(X_d,\ C_d)$. These would have to be explored numerically. Similarly, there may be further bifurcations from the limit cycles in the model of §4.

We have shown in §3.4 that modulating the adaptation rate of the lateral connections according to the output cell variances helps to stabilize the system. However, there is a bifurcation point at the critical value of the coupling constant given in (32). At this bifurcation the kernel of DF_0 has dimension $M(M-1)/2$, where M is the number of output cells. Analysis of this degenerate bifurcation is left for future research.

Lastly, models that perform PCA employ cells with linear post-synaptic response. The role of non-linear response in systems with Hebbian adapta-

tion is almost completely unexplored (however see [20] for some interesting simulation results). We expect that the development of cells with non-linear post-synaptic response (e.g. higher-order nodes) is driven by higher order moments of the input distribution. This extended information may be useful for feature discovery.

Acknowledgements – The author thanks Prof. Dan Hammerstrom and Dr. Bill Baird for lively discussion. David Roe helped automate the bifurcation calculations and Vince Weatherill worked on the figures.

A Stability Calculations

This appendix provides details of the stability calculations of section §3. Recast the equations of motion (14) and (15) in the basis of eigenvectors of Q, writing the projection of ω_i onto e_j as ω_{ij},

$$\dot{\omega}_{ij} = \lambda_j \omega_{ij} + \sum_{m \neq i} \eta_{im} \lambda_j \omega_{mj} - (\sum_m \lambda_m \omega_{im}^2) \omega_{ij} \tag{51}$$

$$\dot{\eta}_{ij} = -d\,(\eta_{ij} + C \sum_m \lambda_m \omega_{im} \omega_{jm}) \quad 1 \leq i,j \leq M. \tag{52}$$

The PCA equilibrium is at $X_0 = \{\omega_{ij} = \delta_{ij}, \eta_{ij} = 0\}$.

The derivatives appearing in the linearization are easily evaluated. First,

$$\left.\frac{\partial \dot{\omega}_{ij}}{\partial \omega_{kl}}\right|_{X_0} = \delta_{ik}\,\delta_{jl}\,(\lambda_j - \lambda_i) - 2\,\delta_{ij}\,\delta_{il}\,\delta_{ik}\,\lambda_i. \tag{53}$$

Next,

$$\left.\frac{\partial \dot{\omega}_{ij}}{\partial \eta_{kl}}\right|_{X_0} = \delta_{ik}\,\delta_{jl}\,\lambda_j, \quad 1 \leq i,j \leq M. \tag{54}$$

The derivatives of the η terms are

$$\left.\frac{\partial \dot{\eta}_{ij}}{\partial \omega_{kl}}\right|_{X_0} = -d\,C\,[\lambda_l\,(\delta_{ik}\,\delta_{jl} + \delta_{jk}\,\delta_{il})] \quad 1 \leq i,j \leq M. \tag{55}$$

and

$$\frac{\partial \dot{\eta}_{ij}}{\partial \eta_{kl}} = -d\,\delta_{ik}\delta_{jl} \quad 1 \leq i,j \leq M. \tag{56}$$

The critical 3×3 sub-blocks of DF_0 are given by the Jacobian matrices

$$\mathcal{M}_{ij} = \frac{\partial(\dot{\omega}_{ij}, \dot{\omega}_{ji}, \dot{\eta}_{ij})}{\partial(\omega_{ij}, \omega_{ji}, \eta_{ij})} = \begin{bmatrix} \lambda_j - \lambda_i & 0 & \lambda_j \\ 0 & \lambda_i - \lambda_j & \lambda_i \\ -d\,C\,\lambda_j & -d\,C\,\lambda_i & -d \end{bmatrix}. \tag{57}$$

The terms in the sub-block $\{A\}$ (22) of DF are from the derivatives of $\dot{\omega}_{ii}$. From (53) the only non-zero terms are

$$\frac{\partial \dot{\omega}_{ii}}{\partial \omega_{ii}} = -2\,\lambda_i. \tag{58}$$

Finally the sub-block (B) (23) contains derivatives of $\dot{\omega}_{iJ}$ with $1 \leq i \leq M$ and $J > M$. From (54) it is clear that all the derivatives with respect to η will vanish. The only remaining terms are

$$\frac{\partial \dot{\omega}_{iJ}}{\partial \omega_{iJ}} = \lambda_J - \lambda_i < 0. \tag{59}$$

Matrix elements outside the block diagonals are easily seen to vanish. This confirms the form of DF_0 given in §3.3. A similar calculation gives the analogous form in §4.2.

The block-diagonal form of DF_0 simplifies the stability calculations since the sub-blocks $\mathcal{M}_{ij}$ are the only parts where an instability can arise. The characteristic polynomial for $\mathcal{M}_{ij}$ is

$$\begin{aligned} P(L) \;=\; L^3 + L^2\, d \;\; &+ \;\; L\,[d\,C(\lambda_i^2+\lambda_j^2) - (\lambda_i-\lambda_j)^2] \\ &+ \;\; d\,(\lambda_i-\lambda_j)^2\,[C(\lambda_i+\lambda_j)-1]. \end{aligned} \tag{60}$$

We applied the Routh-Hurwitz conditions to (60) to derive the stability conditions in §3.3. As a check, it is straightforward to verify that $P(L)$ develops a simple zero root at $C = C_{ij_0} \equiv 1/(\lambda_i+\lambda_j)$. Furthermore $P(L)$ develops a pair of pure imaginary roots

$$\pm i\,(\lambda_i-\lambda_j)\sqrt{C\,(\lambda_i+\lambda_j)-1}$$

at $d = d_{ij_0} \equiv (\lambda_i-\lambda_j)^2(\lambda_i+\lambda_j)/(\lambda_i^2+\lambda_j^2)$.

B Bifurcation Calculations

The bulk of the bifurcation calculations were performed with a symbolic manipulation program, used both interactively and running code written explicitly for this study. Here we sketch the calculations. More details on the Liapunov-Schmidt reduction can be found in Golubitsky and Schaeffer (1984).

The equations of motion for the weights

$$\dot{X} = F(X,\, C),$$

have an equilibrium at X_0. We assume that the stability conditions, (25), are violated for a ***single*** pair of indices (i,j). At $C = C_0$, one of the eigenvalues of $\mathcal{M}_{ij}$ becomes zero, and so the linear part of the vector field, DF_0, has a 1-D kernel. The reduction proceeds as follows. Let S denote the $MN + M(M-1)/2-$ dimensional vector space of variables (ω, η).

1. Split S into $\mathrm{Ker}\,(DF_0)$ and its orthogonal complement, and also into $\mathrm{Range}\,(DF_0)$ and its orthogonal complement. The basis for the kernel and the $(\mathrm{Range})^\perp$ are v_r and v_l, the right and left null eigenvectors of DF_0. Define the projection $E : S \mapsto Range(DF_0)$, and the complementary projection $1 - E$.

2. Points in the configuration space are given coordinates as $X = X_0 + z\,v_r + W$ with $W \in (\mathrm{Ker}\,DF_0)^\perp$ a solution to

$$E\,F(X_0 + z\,v_r + W(z,C), C) = 0. \tag{61}$$

   The solution defines (locally) a 2-D submanifold of S. (Note that $W(0, C_0) = 0$ since $F(X_0, C_0) = 0$ by assumption.)

3. Define the reduced function

$$g(z,C) = (1-E)\,F(X_0 + z\,v_r + W(z,C),C) = v_l \cdot F(X_0 + z\,v_r + W(z,C),C) \quad (62)$$

The zero set of F is in one-one correspondence with the zero set of g. The latter is the bifurcation diagram for the system.

4. The equilibria correspond to $g(z,C) = 0$. If $v_l \cdot v_r > 0$ and $\partial g(z,C)/\partial z < 0$, then the equilibrium corresponding to (z,C) is asymptotically stable, and unstable if $\partial g(z,C)/\partial z > 0$.

In practice (61) is solved for the Taylor series expansion of W. The terms in the series are then substituted into a series expansion of (62).

As an example we carry out the reduction for a network of $M = 2$ cells in a 3-dimensional input space (7 degrees of freedom). Calculations show that the results generalize to nets of arbitrary size. The coordinates are ordered as

$$X = [\omega_{12},\ \omega_{21},\ \eta_{12},\ \omega_{11},\ \omega_{22},\ \omega_{13},\ \omega_{23}],$$

with the equilibrium at

$$X_0 = [0,0,0,1,1,0,0].$$

The linear part of the vector field at X_0 is

$$\mathrm{DF}_0 = \begin{bmatrix} \lambda_2-\lambda_1 & 0 & \lambda_2 & & & & \\ 0 & \lambda_1-\lambda_2 & \lambda_1 & & & & \\ -C\,d\,\lambda_2 & -C\,d\,\lambda_1 & -d & & & & \\ & & & -2\,\lambda_1 & & & \\ & & & & -2\,\lambda_2 & & \\ & & & & & \lambda_3-\lambda_1 & \\ & & & & & & \lambda_3-\lambda_2 \end{bmatrix}. \quad (63)$$

At C_0 the right and left zero eigenvectors of DF_0 are given by

$$v_r = [\frac{\lambda_2}{(\lambda_1-\lambda_2)},\ \frac{\lambda_1}{(\lambda_2-\lambda_1)},\ 1,\ 0,\ 0,\ 0,\ 0] \quad (64)$$

and

$$v_l = -[\frac{d\lambda_2}{(\lambda_1+\lambda_2)(\lambda_2-\lambda_1)},\ \frac{d\lambda_1}{(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)},\ 1,\ 0,\ 0,\ 0,\ 0]. \quad (65)$$

The coefficients in the series expansion for W are found by differentiating (61) and solving the resulting expression for the coefficients. The required coefficients are

$$W_z = -DF^{-1}\,E\,DF_0[v_r] = 0 \quad (66)$$

$$W_c = -DF^{-1}\,E\,F_C(X_0,C_0) \quad (67)$$

$$W_{zz} = -DF^{-1}\,E\,D^2F_0[v_r,v_r] \quad (68)$$

where DF^{-1} is the inverse of DF_0 restricted to Range(DF_0). (DF_0 is an isomorphism from $(KerDF_0)^{\perp}$ to Range(DF_0).) In the present case, $F_C(X_0, C_0)$ vanishes so

$$W_c = [0, 0, 0, 0, 0, 0, 0]. \tag{69}$$

Moving on to W_{zz},

$$\mathrm{D^2F_0[v_r, v_r]} = [0, 0, 0, \frac{-2\,(\lambda_1^3 - \lambda_1^2\lambda_2 + \lambda_2^3)}{(\lambda_1 - \lambda_2)^2}, \frac{-2\,(\lambda_1^3 - \lambda_1\lambda_2^2 + \lambda_2^3)}{(\lambda_1 - \lambda_2)^2}, 0, 0]. \tag{70}$$

This is already perpendicular to v_l, so the action of E is the identity. Further the preimage of $D^2F_0[v_r, v_r]$ under DF_0 is trivial and we have

$$W_{zz} = [0, 0, 0, \frac{-\lambda_1^3 + \lambda_1^2\lambda_2 - \lambda_2^3}{\lambda_1(\lambda_1 - \lambda_2)^2}, \frac{-\lambda_1^3 + \lambda_1\lambda_2^2 - \lambda_2^3}{\lambda_2(\lambda_1 - \lambda_2)^2}, 0, 0] \tag{71}$$

The terms in the series expansion of the reduced function needed to identify the bifurcation are found by differentiating (62),

$$\begin{aligned}
g_z &= v_l \cdot DF_0[v_r + W_z] = 0 \text{ using (66).} \\
g_c &= v_l \cdot F_c(X_0, C_0) && (72) \\
g_{zz} &= v_l \cdot D^2F_0[v_r, v_r] && (73) \\
g_{zc} &= v_l \cdot (DF_c \cdot v_r + D^2F_0[v_r, W_c]) && (74) \\
g_{zzz} &= v_l \cdot (D^3F_0[v_r, v_r, v_r] + 3\,D^2F_0[v_r, W_{zz}]) && (75)
\end{aligned}$$

To find g_{zc} we note that the only non-zero elements of the matrix DF_c are

$$\begin{aligned}
(DF_c)_{3,1} &= -d\,\lambda_2 \\
(DF_c)_{3,2} &= -d\,\lambda_1. && (76)
\end{aligned}$$

Finally, the only non-zero terms in the tensor D^3F_0 arise from the cubic terms in F. We find

$$\begin{aligned}
g_c &= 0 \\
g_{zz} &= 0 \quad \text{using (65) and (70) in (73).} \\
g_{zc} &= -d(\lambda_1 + \lambda_2) \quad \text{using (65), (67), (76), and in (74).} \\
g_{zzz} &= 6\,d.
\end{aligned}$$

To third order, the reduced function is thus

$$\begin{aligned}
g &= g_{cz}(C - C_0)\,z + \frac{1}{3!}g_{zzz}z^3 + \ldots \\
&= -d(\lambda_1 + \lambda_2)\,(C - C_0)\,z \;+\; d\,z^3. && (77)
\end{aligned}$$

For $C > C_0$, the roots of g are at $z_s = \pm\sqrt{(C - C_0)\,(\lambda_1 + \lambda_2)}$, and $z = 0$. For $C < C_0$ the only root of g is at $z = 0$. Since X_0 corresponds to $z = 0$, the latter is stable for

$C > C_0$, and unstable for $C < C_0$. By exchange of stability, the secondary equilibria z_s should be unstable. As a check we note that for $d > d_0$, $v_r \cdot v_l > 0$. Then,

$$\partial g/\partial z = 3d\, z^2 \; - \; (C - C_0)\, d\, (\lambda_1 + \lambda_2).$$

At $z = 0$ we have $\partial g/\partial z < 0$ for $C > C_0$ indicating stability of X_0. At z_s, we have $\partial g/\partial z > 0$ indicating instability.

References

[1] R. Perez, L. Glass, and R. Shlaer. Development of specificity in the cat visual cortex. *Journal of Mathematical Biology*, 1:275–288, 1975.

[2] L.N. Cooper, F. Liverman, and E. Oja. A theory for the acquisition and loss of neuron specificity in the visual cortex. *Biol. Cybern.*, 33:9–28, 1979.

[3] C. von der Malsburg and J.D. Cowan. Outline of a theory for the ontogenesis of iso-orientation domains in visual cortex. *Biol. Cyb.*, 45:49–56, 1982.

[4] H.G. Barrow. Learning receptive fields. *Proceeding of the IJCNN*, IV:115–121, 1987.

[5] E. Oja. A simplified neuron model as a principal component analyzer. *J. Math. Biology*, 15:267–273, 1982.

[6] Ralph Linsker. Self organization in a perceptual network. *Computer*, pages 105–117, March 1988.

[7] E. Oja and J. Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *J. of Math. Anal. and Appl.*, 106:69–84, 1985.

[8] E. Oja. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1:61–68, 1989.

[9] T. Sanger. An optimality principle for unsupervised learning. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems 1*. Morgan Kauffmann, 1989.

[10] T. Leen, M. Rudnick, and D. Hammerstrom. Hebbian feature discovery improves classifier efficiency. In *Proceedings of the IEEE/INNS Internationl Joint Conference on Neural Networks*, pages I–51 – I–56, June 1990.

[11] P. Foldiak. Adaptive network for optimal linear feature extraction. In *Proceedings of the IJCNN*, pages I 401–405, 1989.

[12] J. Rubner and K. Schulten. Development of feature detectors by self-organization: A network model. *Biol. Cyb.*, 62:193–199, 1990.

[13] T. Kohonen. *Self-Organization and Associative Memory, 2nd Edition.* Springer-Verlag, Berlin, 1988.

[14] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control*, 22:551–575, 1977.

[15] M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, Inc., San Diego, 1974.

[16] A.L. Yuille, D.M. Kammen, and D.S. Cohen. Quadrature and the development of orientation selective cortical cells by hebb rules. *Biol. Cybern.*, 61:183–194, 1989.

[17] A. Fuchs and H. Haken. Pattern recognition and associative memeory as dynamical processes in a synergetic system. *Biol. Cybern.*, 60:17–22, 1988.

[18] Martin Golubitsky and David Schaeffer. *Singularities and Groups in Bifurcation Theory, Vol. I.* Springer-Verlag, New York, 1984.

[19] Morris W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2:331–349, 1989.

[20] Arthur Carlson. Anti-hebbian learning in a non-linear neural network. *Biol. Cyb.*, 64:171–176, 1990.

[21] J. Marsden and M. McCracken. *The Hopf Bifurcation and Its Applications.* Springer-Verlag, New York, 1976.

Figure Captions

1. Bifurcation diagram for the minimal model. Heavy lines are stable branches and light lines are unstable branches. Insets show the receptive fields corresonding to equilibria on the stable branches.

2. Hysteresis curve obtained by tracing out the bifurcation diagram of Fig. 1 for a 2-cell model.

3. Receptive fields generated in a 3-cell simulation.
   a) Weight vectors at $C = 1.0$ are the eigenvectors of the autocorrelation.
   b) At $C = 0.28$, $\omega_1 = e_1$, $\omega_2 = 0.477\ e_3 - 0.768\ e_2$,
   $\omega_3 = 0.477\ e_3 + 0.768\ e_2$. c) At $C = 0.18$, $\omega_1 = \omega_3 = 0.766\ e_1$, $\omega_2 = 0.999\ e_2$.

4. Regions in the $(\lambda_1,\ \lambda_2)$ plane corresponding to super-critical (shaded) and sub-critical (unshaded) Hopf bifurcations in the complete model.

5. Stable oscillations and equilibria near a super-critical Hopf bifurcation in the complete model. See text for explanation.

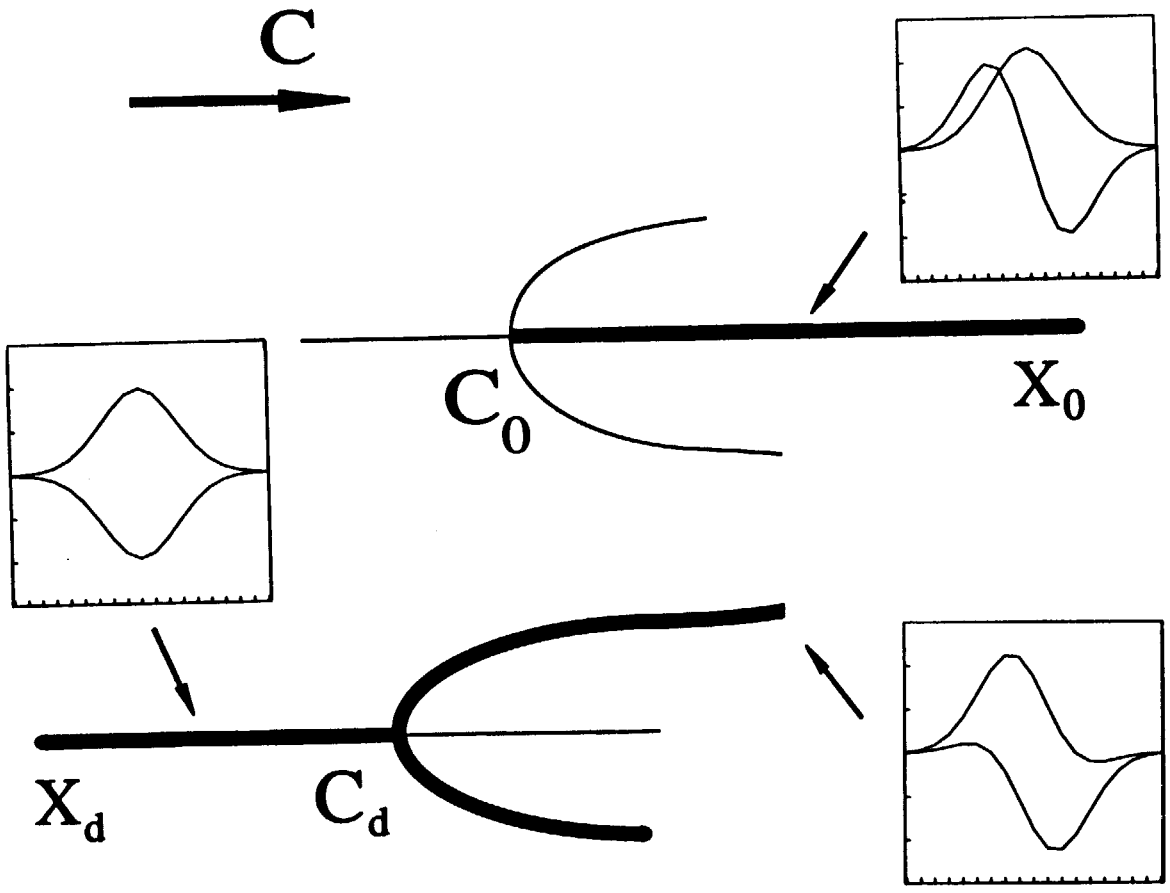

FIGURE 1.

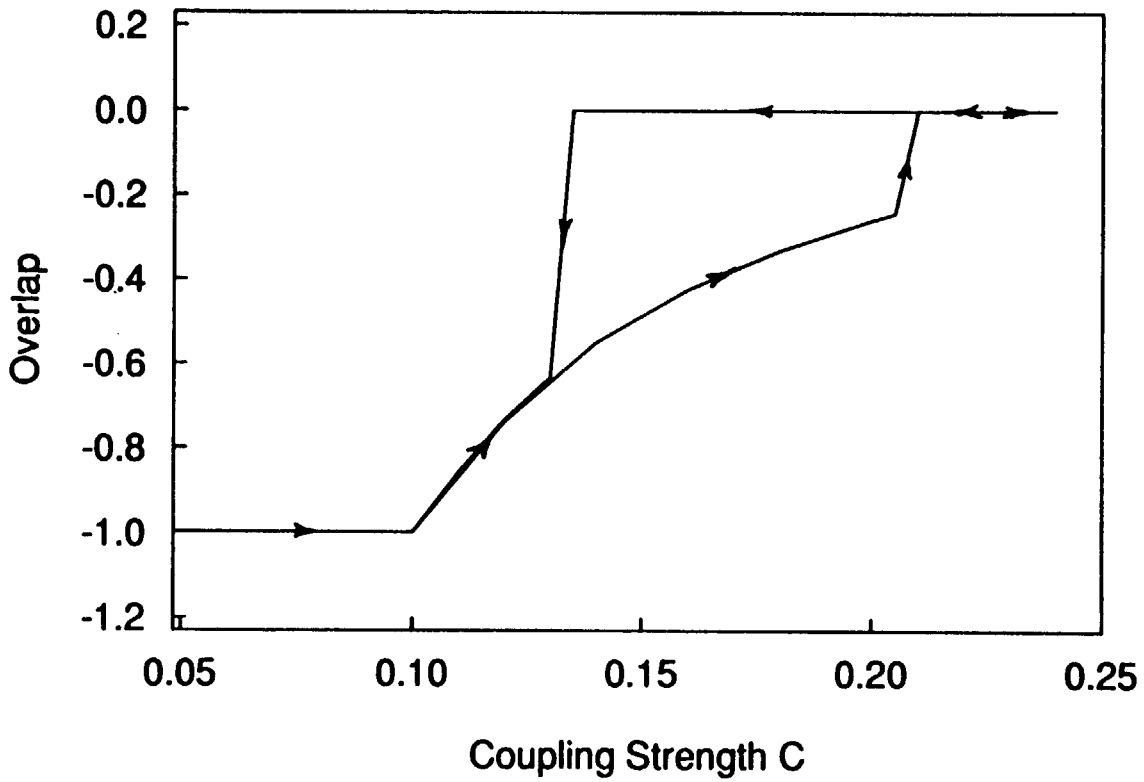
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
Overlap
0.05
0.10
0.15
0.20
0.25
Coupling Strength C

FIGURE 2.

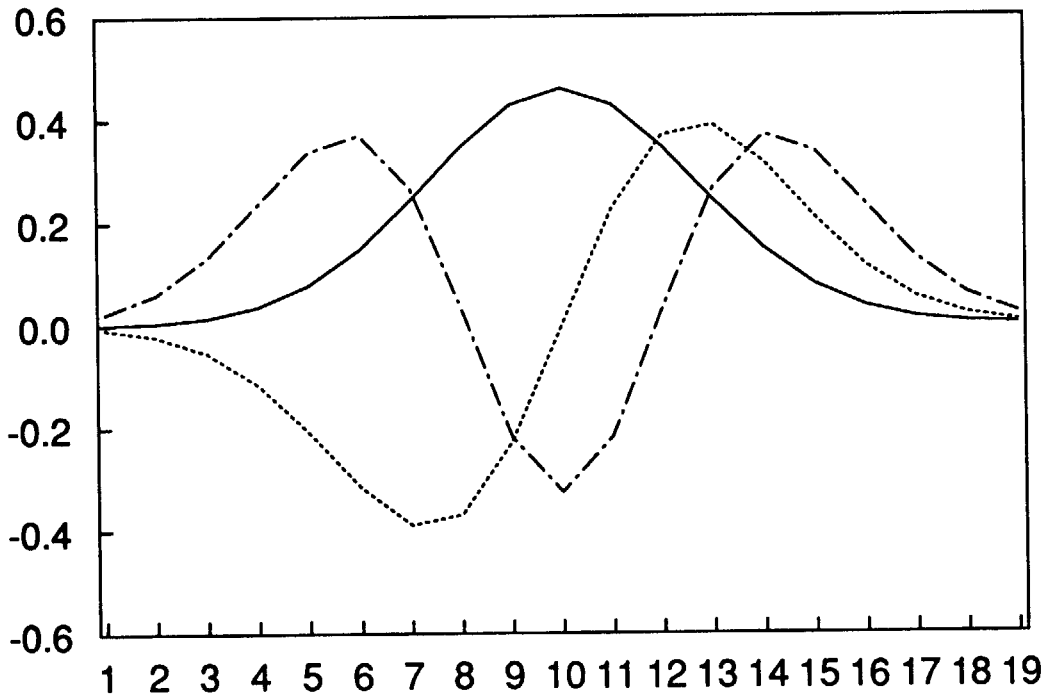

FIGURE 3(a).

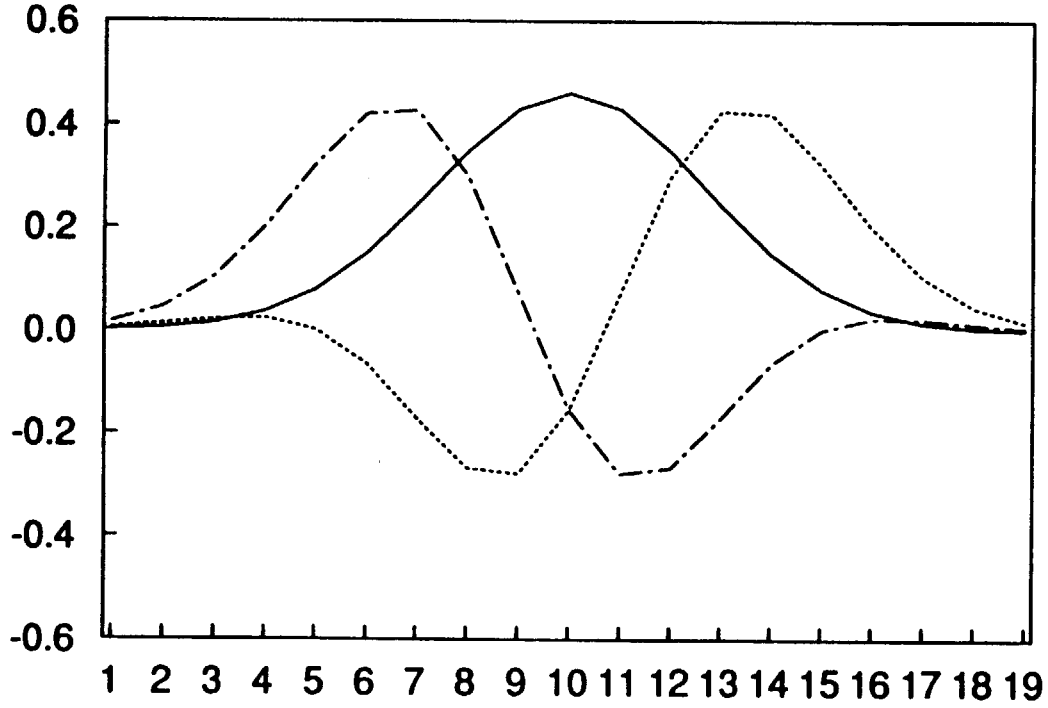

FIGURE 3(b).

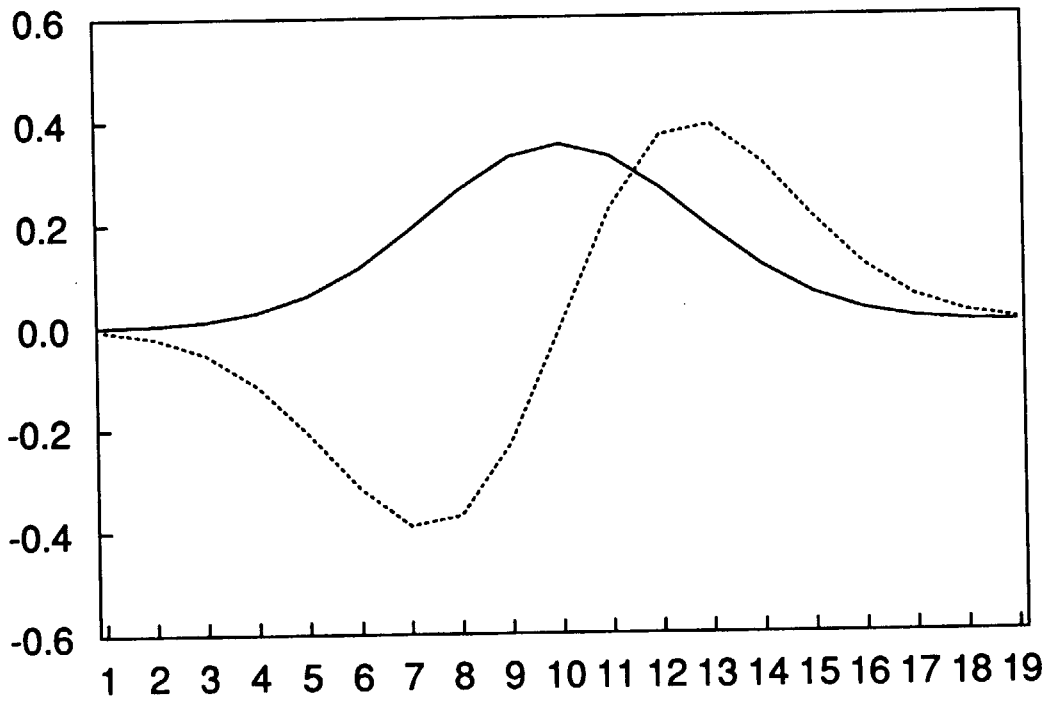

FIGURE 3(c).

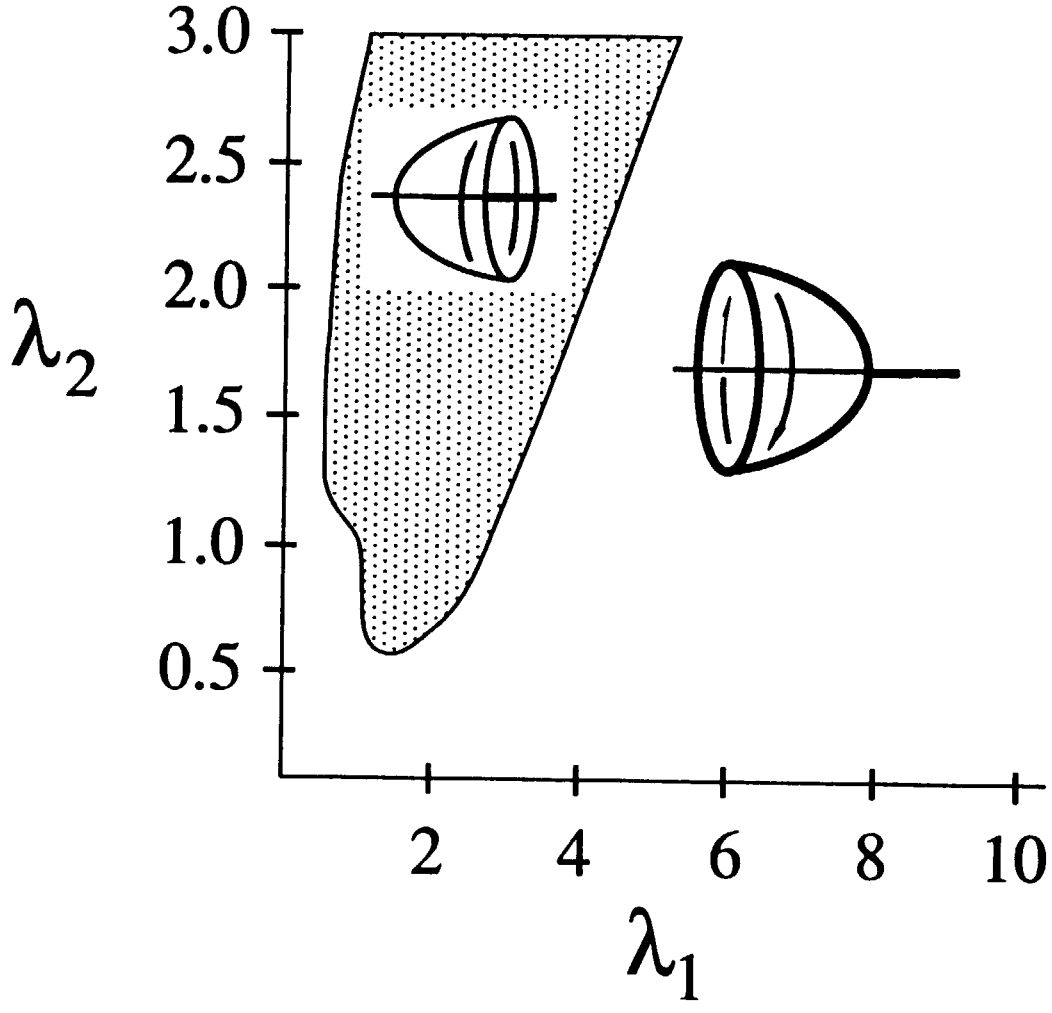

FIGURE 4.

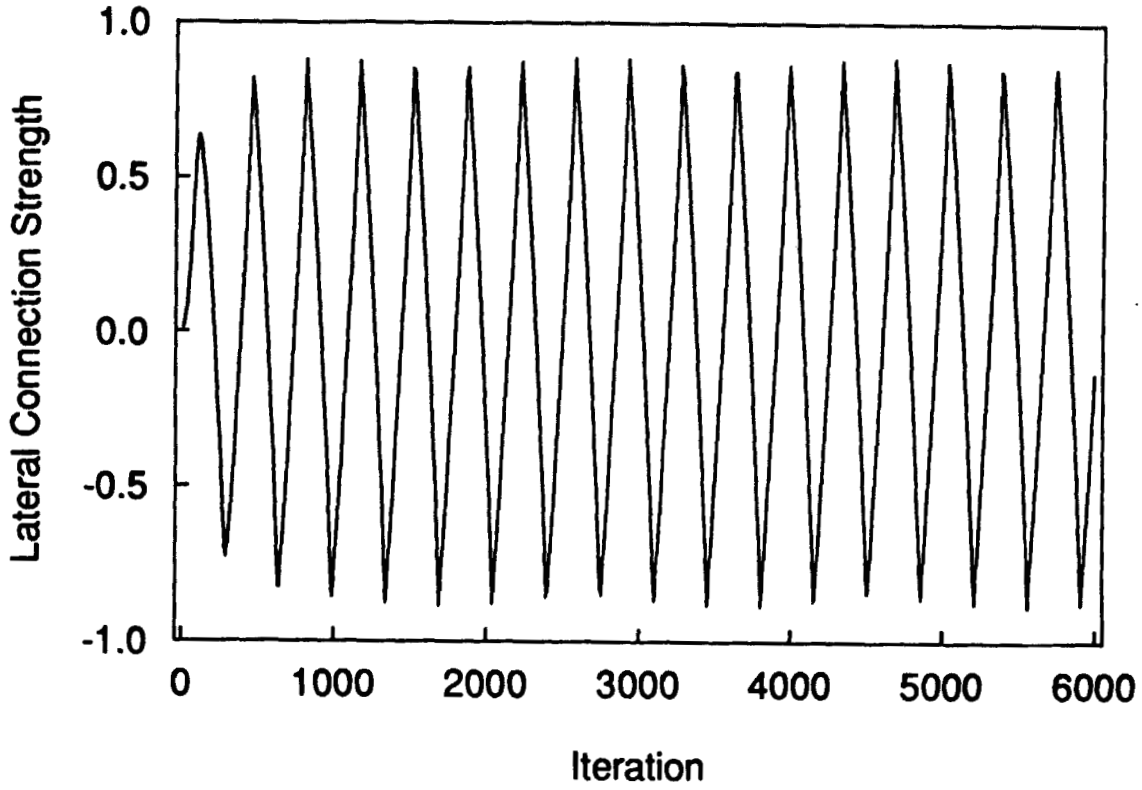
1.0
0.5
0.0
-0.5
-1.0
Lateral Connection Strength
0
1000
2000
3000
4000
5000
6000
Iteration

FIGURE 5(a).

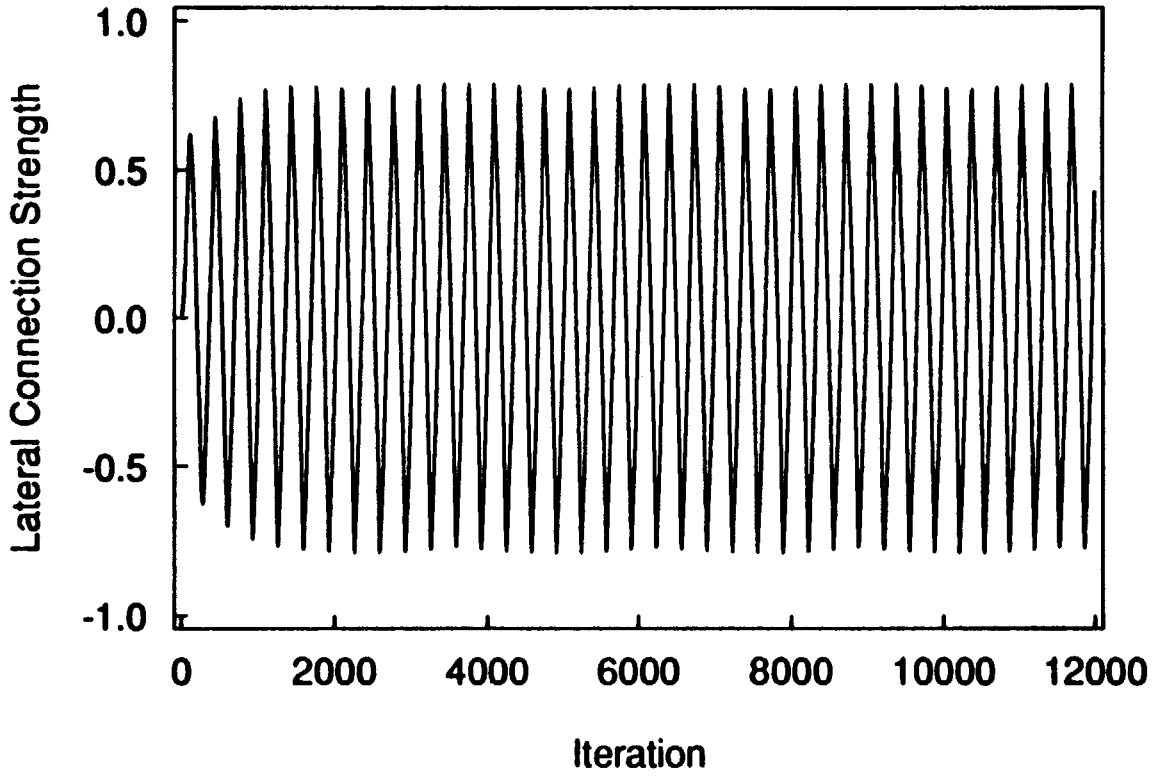
1.0
0.5
0.0
-0.5
-1.0
Lateral Connection Strength
0
2000
4000
6000
8000
10000
12000
Iteration

FIGURE 5(b).

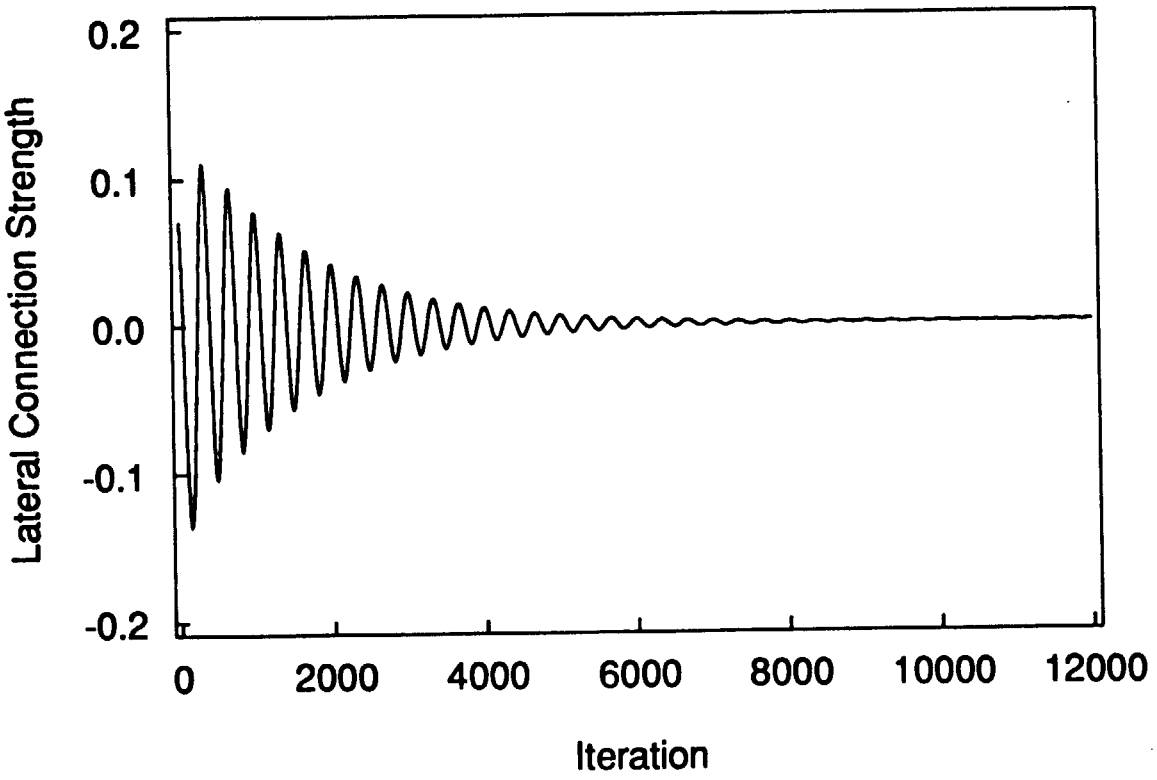
0.2
0.1
0.0
-0.1
-0.2
Lateral Connection Strength
0
2000
4000
6000
8000
10000
12000
Iteration

FIGURE 5(c).

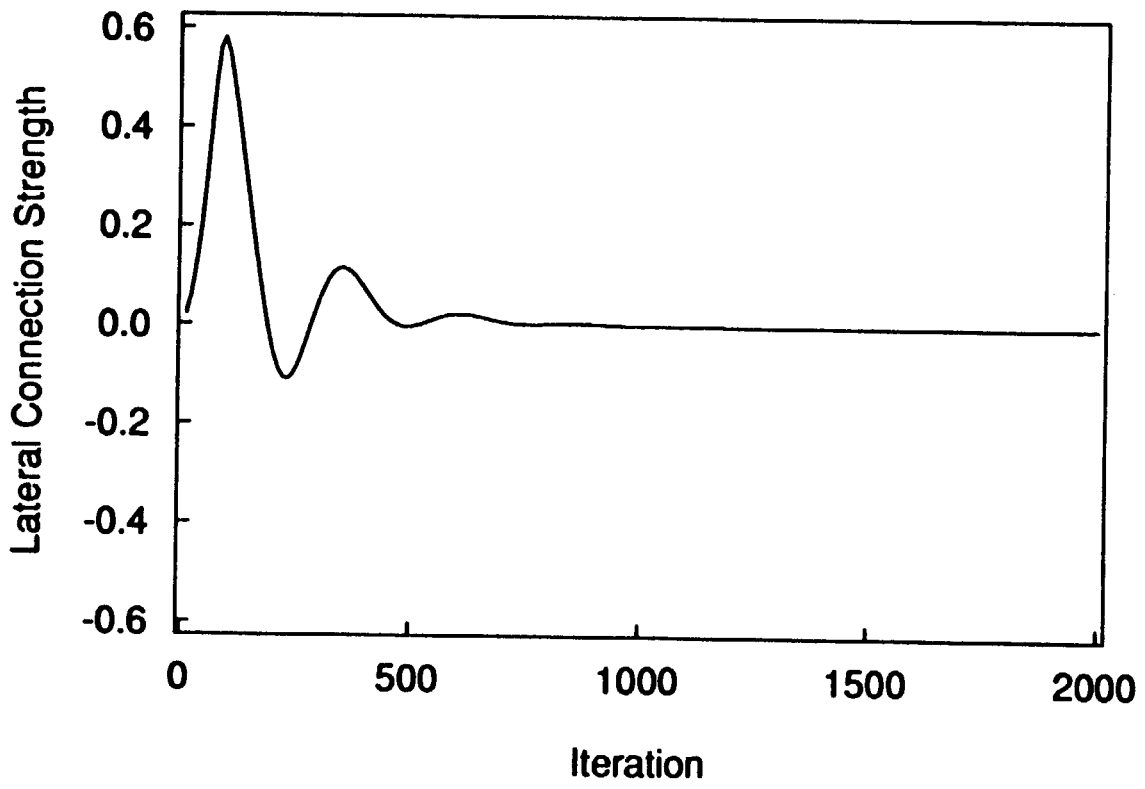
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
Lateral Connection Strength
0
500
1000
1500
2000
Iteration

FIGURE 5(d).